**Objetivo**

Este trabalho tem como objetivo ajudar na resposta de perguntas referentes ao impacto que o Covid19 ocasionou no mundo, através de informações como mortes ocasionadas pelo Covid19, número de pessoas que conseguiram se curar e números de novas pessoas infectadas.

Perguntas que desejo responder:

* Qual os top 10 países que tiveram o maior número de novos casos?
* Qual o período com o maior número de novos casos?
* Qual a proporção de novos casos de 19, mortes por Covid19 e recuperação do Covid19 durante um período?
* Qual a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 10 países?
* Qual o comportamento de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo?
* Qual o número de novos casos por continente?
* Qual a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 no Brasil?
* Qual o número de pessoas recuperadas que tomaram a vacina contra Covid19?

**Coleta**

Para conseguir cumprir o objetivo deste trabalho foi necessário buscar um dataset que tivesse os dados para responder as perguntas, através do site <https://www.kaggle.com/> consegui o dataset que me ajudou com as informações.

Utilizei o dataset <https://www.kaggle.com/datasets/imdevskp/corona-virus-report?select=full_grouped.csv>.

Depois de obter o dataset fiz a importação dele no <https://community.cloud.databricks.com> através da opção DBFS.

Como usei a opção DBFS:

1. Cliquei a opção catalog
2. Apertei na aba DBFS
3. Apertei em UPLOAD
4. Selecionei o arquivo que estava no meu computador.

**Modelagem**

Para o modelagem utilizei um CSV extraído do <https://www.kaggle.com/>, fazendo o upload desse CSV no <https://community.cloud.databricks.com>, depois de fazer o upload utilizei Python para criar uma tabela com os dados desse csv no databricks.

As outras tabelas foram populadas e crias através do SQL

**Dados armazenados no CSV:**

* Lista de Países
* Lista de Continentes
* Data
* Número de mortes
* Número de pessoas recuperadas
* Número de pessoas infectadas
* Número de pessoas ainda infectadas
* Número de novos casos de Covid19
* Número de novas mortes
* Número de novas pessoas recuperadas

**Período dos dados :** 21/01/2020 até 26/07/2020

**Catálogo de dados:**

| tbGoldCovid19Dataset | | |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Data | STRING | Data que foi coletado o dado informado |
| País | STRING | País onde foi coletado o dado |
| Confirmados | STRING | Número de pessoas infectadas |
| Mortes | STRING | Número de mortes |
| Recuperados | STRING | Número de pessoas recuperadas |
| Ativos | STRING | Número de pessoas ainda infectadas |
| Novos\_casos | STRING | Número de novos casos de Covid19 |
| Novas\_mortes | STRING | Número de novas mortes |
| Novos\_recuperados | STRING | Número de novas pessoas recuperadas |
| Continente | STRING | Continente onde foi coletado o dado |

| tbSilverCovid19Dataset | | |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Data | STRING | Data que foi coletado o dado informado |
| País | STRING | País onde foi coletado o dado |
| Confirmados | STRING | Número de pessoas infectadas |
| Mortes | STRING | Número de mortes |
| Recuperados | STRING | Número de pessoas recuperadas |
| Ativos | STRING | Número de pessoas ainda infectadas |
| Novos\_casos | STRING | Número de novos casos de Covid19 |
| Novas\_mortes | STRING | Número de novas mortes |
| Novos\_recuperados | STRING | Número de novas pessoas recuperadas |
| Continente | STRING | Continente onde foi coletado o dado |

| full\_grouped\_base\_dataset\_csv | | |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Date | STRING | Data que foi coletado o dado informado |
| Country\_Region | STRING | País onde foi coletado o dado |
| Confirmed | STRING | Número de pessoas infectadas |
| Deaths | STRING | Número de mortes |
| Recovered | STRING | Número de pessoas recuperadas |
| Active | STRING | Número de pessoas ainda infectadas |
| New\_cases | STRING | Número de novos casos de Covid19 |
| New\_deaths | STRING | Número de novas mortes |
| New\_recovered | STRING | Número de novas pessoas recuperadas |
| WHO\_Region | STRING | Continente onde foi coletado o dado |

| De Para de mudança de nomenclatura | | | |
| --- | --- | --- | --- |
| Tabela | Coluna com nome antes | Tabela | Coluna com nome depois |
| full\_grouped\_base\_dataset\_csv | Date | tbSilverCovid19Dataset | Data |
| full\_grouped\_base\_dataset\_csv | Country\_Region | tbSilverCovid19Dataset | País |
| full\_grouped\_base\_dataset\_csv | Confirmed | tbSilverCovid19Dataset | Confirmados |
| full\_grouped\_base\_dataset\_csv | Deaths | tbSilverCovid19Dataset | Mortes |
| full\_grouped\_base\_dataset\_csv | Recovered | tbSilverCovid19Dataset | Recuperados |
| full\_grouped\_base\_dataset\_csv | Active | tbSilverCovid19Dataset | Ativos |
| full\_grouped\_base\_dataset\_csv | New\_cases | tbSilverCovid19Dataset | Novos\_casos |
| full\_grouped\_base\_dataset\_csv | New\_deaths | tbSilverCovid19Dataset | Novas\_mortes |
| full\_grouped\_base\_dataset\_csv | New\_recovered | tbSilverCovid19Dataset | Novos\_recuperados |
| full\_grouped\_base\_dataset\_csv | WHO\_Region | tbSilverCovid19Dataset | Continente |

**Carga**

A primeira carga foi feita através de um CSV extraído do <https://www.kaggle.com/>, fazendo o upload desse CSV no <https://community.cloud.databricks.com>, depois de fazer o upload utilizei Python para criar uma tabela com os dados desse csv no databricks.

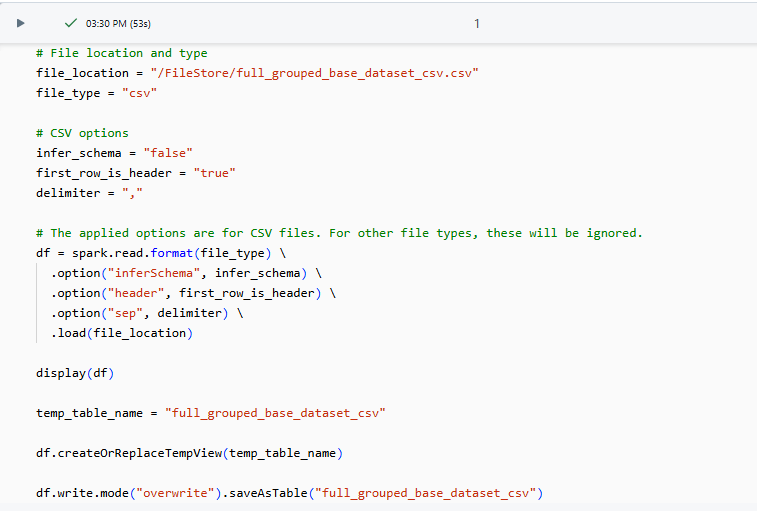


Imagem:Criação da tabela full\_grouped\_base\_dataset\_csv com os dados do csv.

A carga da tabela tbSilverCovid19Dataset foi feita através de um “insert select” da tabela full\_grouped\_base\_dataset\_csv originada do CSV.



Imagem: Carga da tabela tbSilverCovid19Dataset

A carga da tabela tbGoldCovid19Dataset foi feita através de um “insert select” da tabela tbSilverCovid19Dataset utilizando um filtro para tirar informações que só continham valor 0 nas colunas de número.

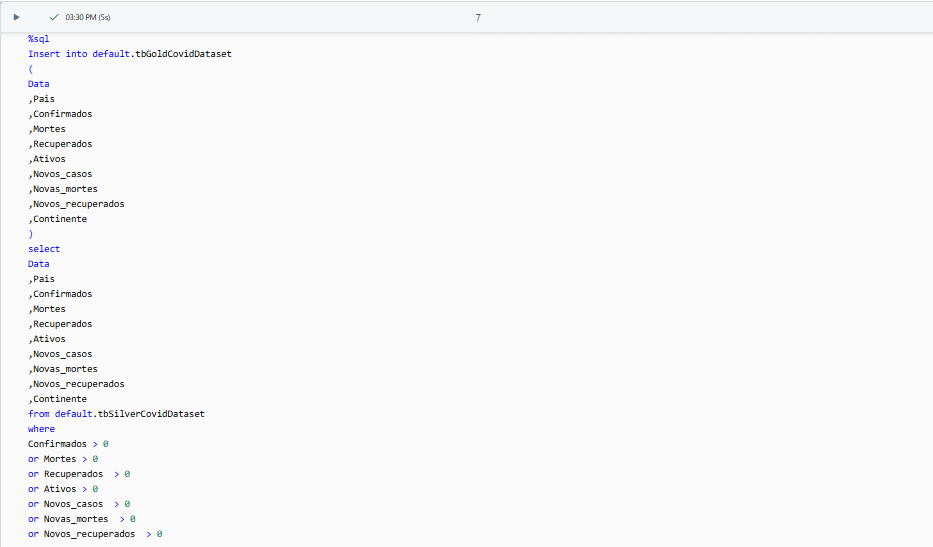


Imagem: Carga da tabela tbGoldCovid19Dataset

**Análise**

Após analisar os dados do csv não obtive problema com a qualidade dos dados, as colunas possuíam corretamente suas respectivas informações, por exemplo, coluna de país mostravam países, colunas de números mostravam números.

Depois de toda a modelagem e o carregamento das tabelas iniciei a análise dos dados para responder meu objetivo, respondendo às seguintes perguntas:

**Qual os top 10 países que tiveram o maior número de novos casos ?**

**Resposta**:US, Brasil, Índia, Rússia, Peru, Chile, South America, Mexico, United Kingdom e Iran.

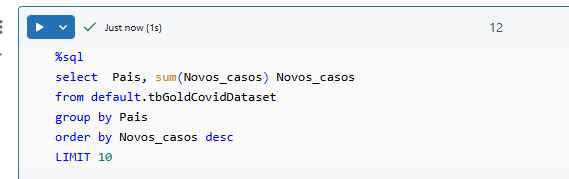


Imagem: Query que trás os top 10 países com o maior número de novos casos.



Imagem: Resultado da query que trás os top 10 países com o maior número de novos casos.

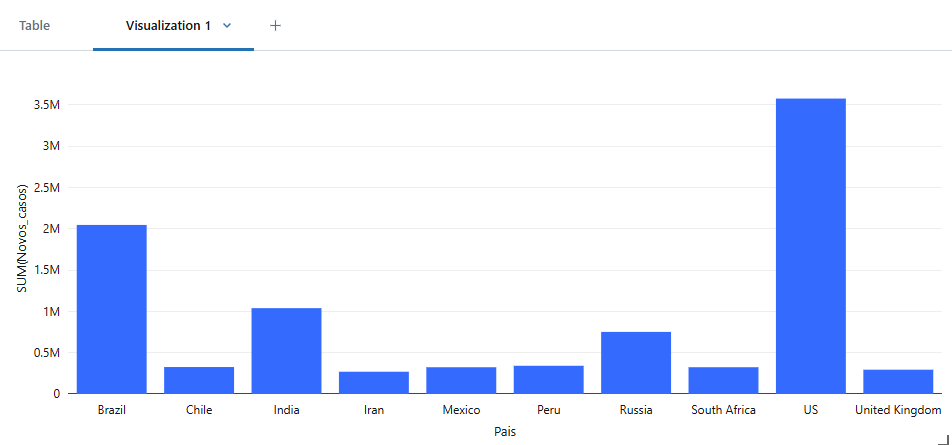


Imagem: Gráfico do resultado da query que trás os top 10 países com o maior número de novos casos.

**Qual o período com o maior número de novos casos?**

**Resposta:** o período com maior número de novos casos é em Junho de 2025.

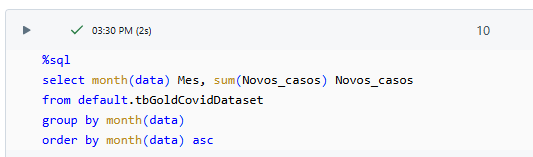


Imagem: Query que trás o período com o maior número de novos casos.

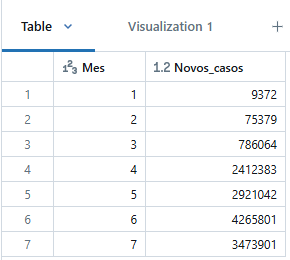


Imagem: resultado que trás o período com o maior número de novos casos.

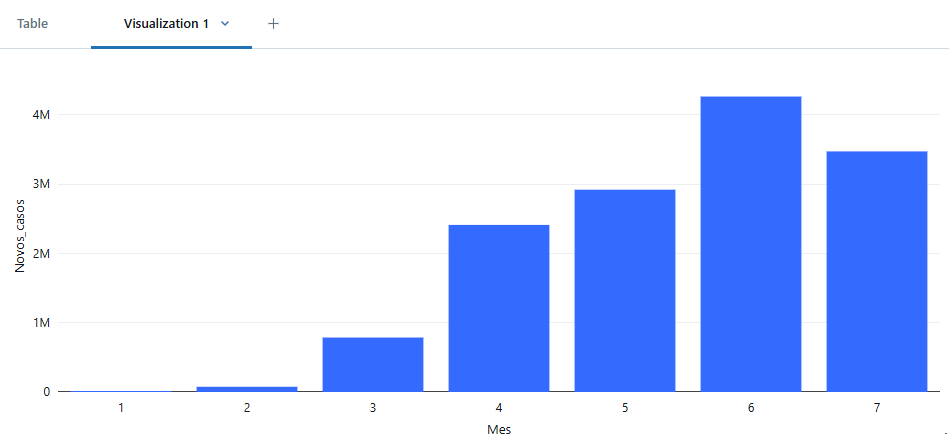


Imagem: Gráfico da query que trás o período com o maior número de novos casos.

**Qual a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 durante um período?**

**Resposta:**Depois de analisar podemos ver que conforme o período foi passando o número de pessoas recuperada foi aumentando mais que o número de novos casos.

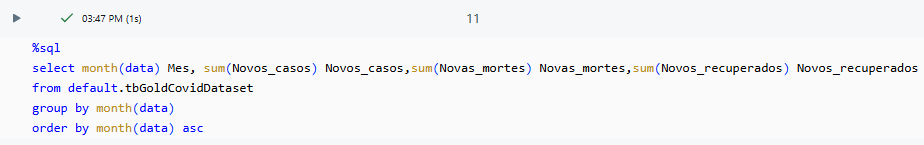


Imagem: Query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 durante um período.

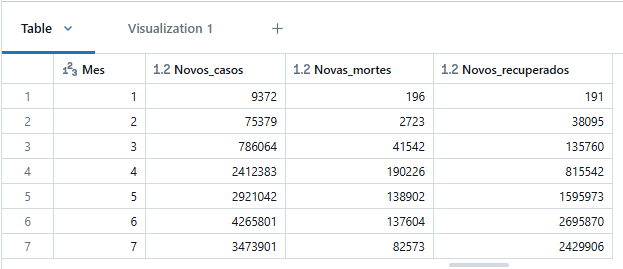


Imagem: Resultado da query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 durante um período.

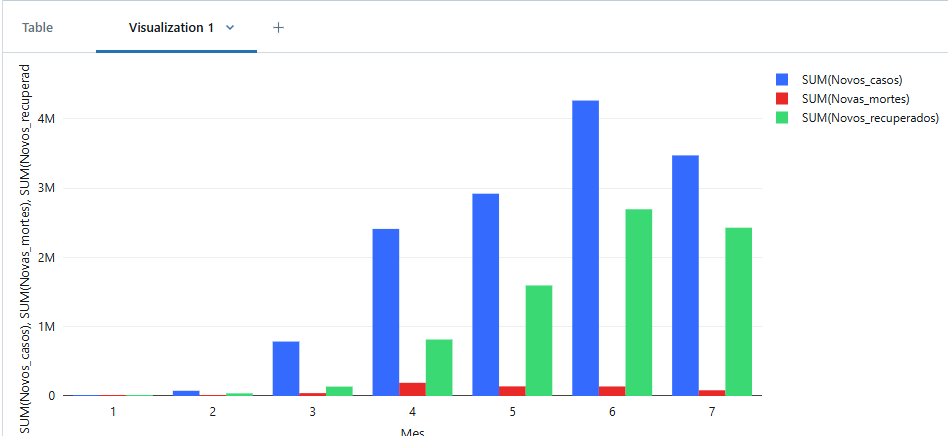


Imagem: Gráfico do resultado da query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 durante um período.

**Qual a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 10 países?**

**Resposta:** Depois de analisar podemos ver que em alguns países o número de novos casos foi muito maior que o número de pessoas recuperadas, enquanto em outros os valores ficaram um pouco mais próximos.

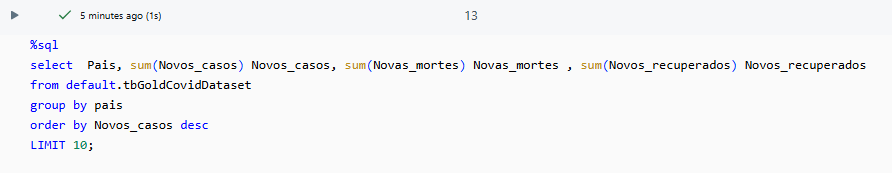


Imagem:Query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 10 países.

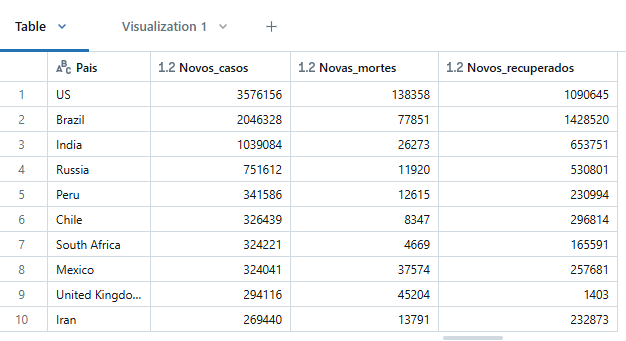


Imagem:Resultado da query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 10 países.

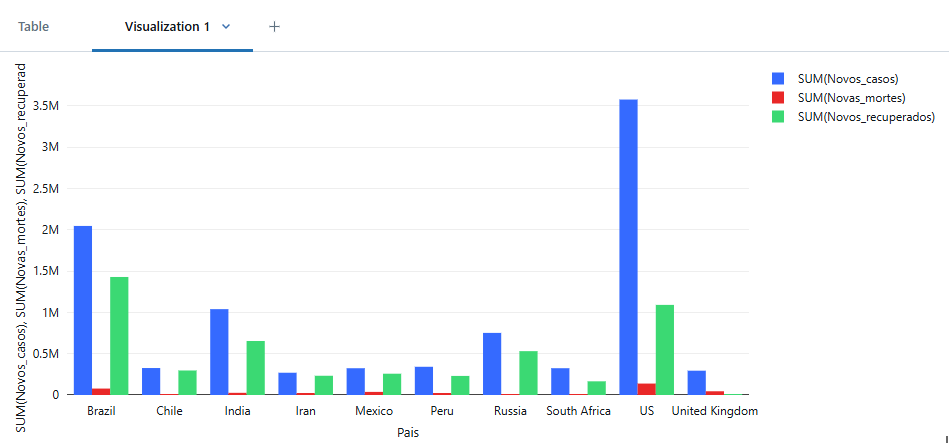


Imagem:Gráfico do resultado da query da proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 10 países.

**Qual o comportamento de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo?**

**Resposta:** Depois de analisar podemos ver que os os países menos os US conforme foi passando o período o número de pessoas recuperados foi chegando mais próximo do número de novos casos.

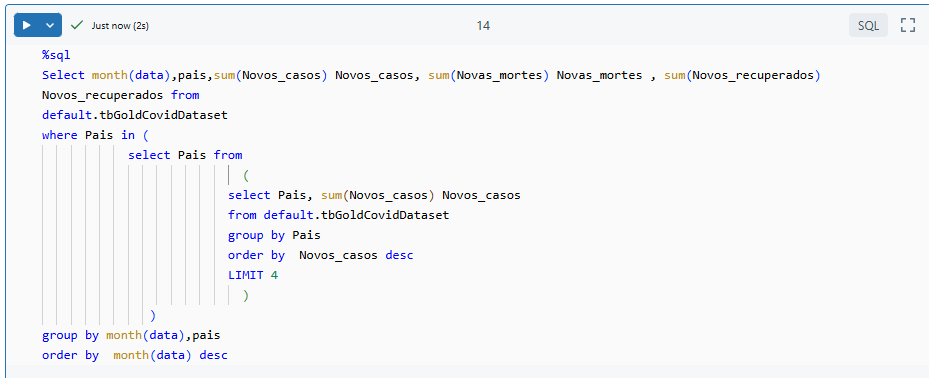


Imagem: Query do comportamento de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo.



Imagem: Resultado da query do comportamento de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo (Parte 1).

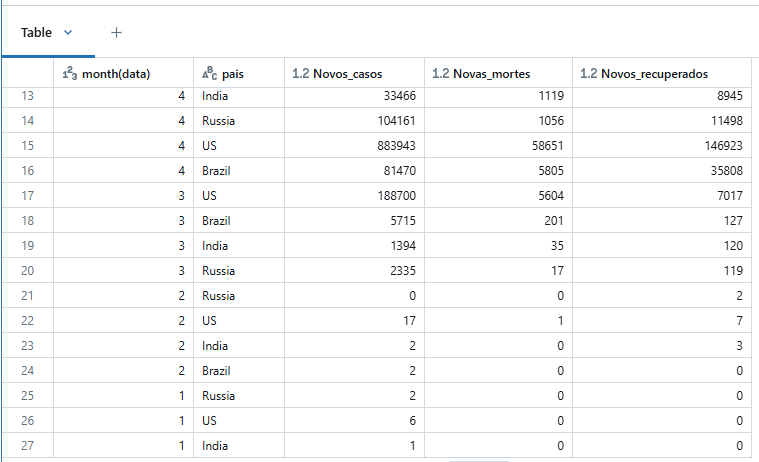


Imagem: Resultado da query do comportamento de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo (Parte 2).

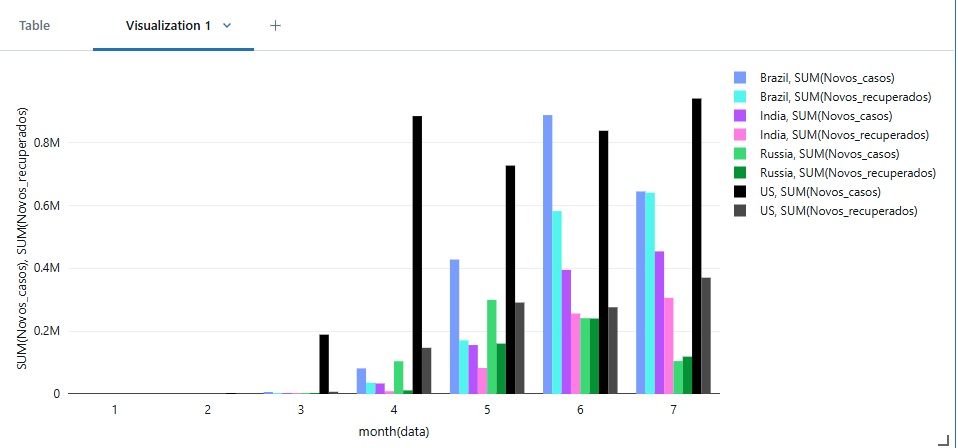


Imagem: Gráfico do resultado da query do comportamento de novos casos de Covid19 e recuperação do Covid19 dos top 4 países com maiores números de novos casos por em um período de tempo.

**Qual o número de novos casos por continente?**

**Resposta:** Depois de analisar podemos ver que o continente com o maior número de novos casos foi a América.

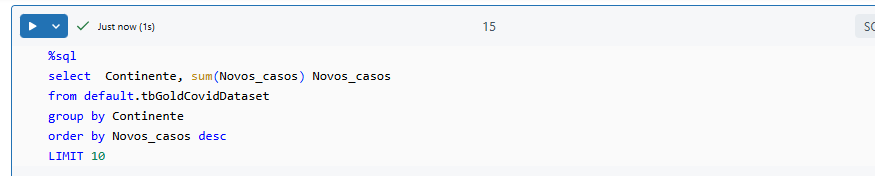


Imagem: Query do número de novos casos por continente.

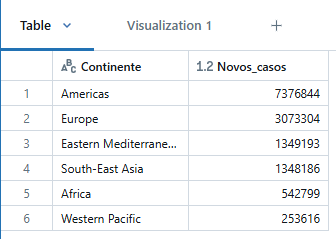


Imagem: Resultado da query do número de novos casos por continente.

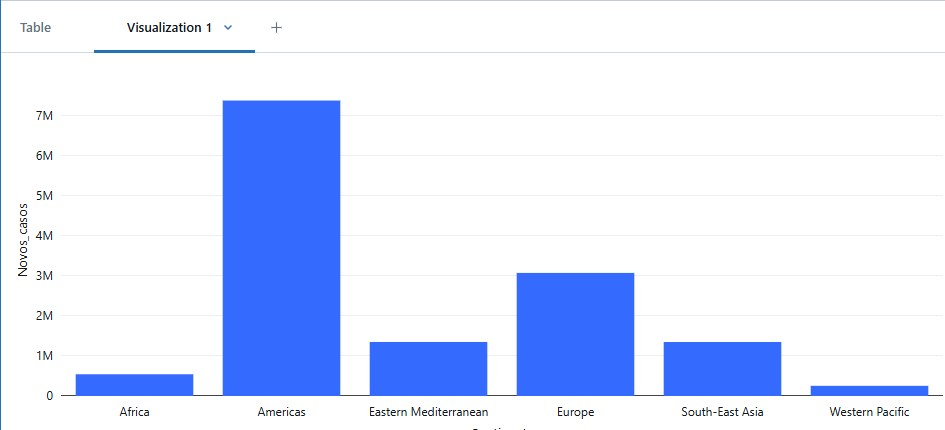


Imagem: Gráfico do resultado da query do número de novos casos por continente.

**Qual a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 no Brasil?**

**Resposta:** Depois de analisar podemos ver que no início do período o teve mais mais novos casos do que pessoas recuperadas, no final do período os valores quase se igualaram.

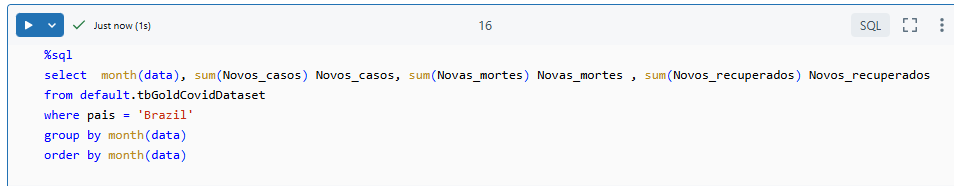


Imagem: Query a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 no Brasil.

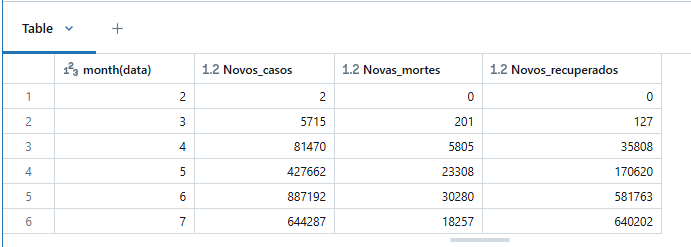


Imagem: Resultado da query a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 no

Brasil.

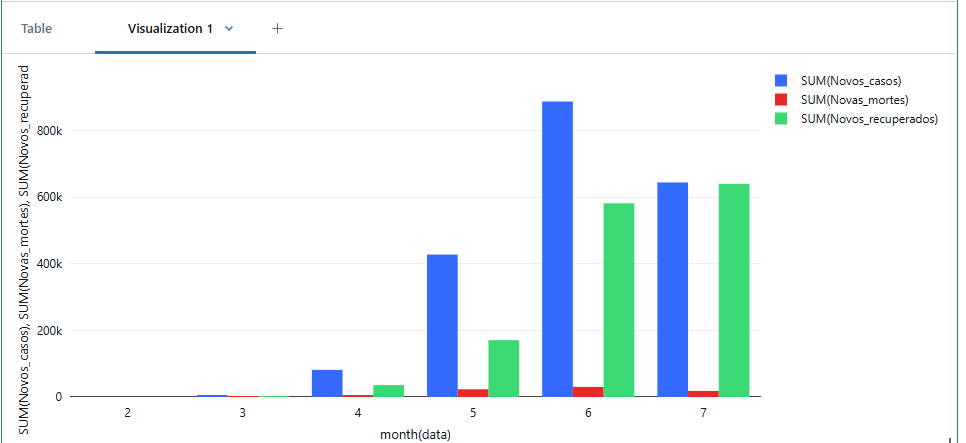


Imagem: Gráfico do resultado da query a proporção de novos casos de Covid19, mortes por Covid19 e recuperação do Covid19 no Brasil.

**Autoavaliação**

Infelizmente não consegui responder minhas perguntas completamente, meus dados só tinham informações de 21/01/2020 até 26/07/2020, um período muito curto para fazer uma análise completa do Covid19 no mundo, não possuía o primeiro e último mês completos podendo mascarar uma análise dos meses, como por exemplo a pergunta “Qual o período com o maior número de novos casos?” pelo gráfico mostrou que foi junho mas será que se julho estivesse com mês fechado seria junho.

Meu dataset não possuía informações de vacinados e não consegui responder a pergunta “Qual o número de pessoas recuperadas que tomaram a vacina contra Covid19?”